# Biến ngẫu nhiên liên tục (Continuos Random Variables)

## Thống kê máy tính và ứng dụng (ĐTTX)

Vũ Quốc Hoàng (vqhoang@fit.hcmus.edu.vn)
FIT - HCMUS

Ngày 3 tháng 4 năm 2022

# Nội dung

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

2. Hàm phân phối

3. Kì vọng và phương sai

4. Một số phân phối xác suất liên tục điển hình

5. Phương pháp xấp xỉ phân phối bằng mô phỏng

# Nội dung

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

2. Hàm phân phối

3. Kì vọng và phương sai

4. Một số phân phối xác suất liên tục điển hình

5. Phương pháp xấp xỉ phân phối bằng mô phỏng

# Biến ngẫu nhiên liên tục và hàm mật độ xác suất

$X$ được gọi là **biến ngẫu liên tục** (continuous random variable) nếu có hàm số không âm $f : \mathbb{R} \to \mathbb{R}$ sao cho với mọi khoảng $[a, b]$ trong $\mathbb{R}$ ta có

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

- $f$ được gọi là **hàm mật độ xác suất** (probability denstity function) của $X$ vì nó cho biết khả năng $X$ nhận giá trị trong các khoảng rất nhỏ của trục số thực $\mathbb{R}$

$$P(a \leq X \leq a + \epsilon) = \int_a^{a+\epsilon} f(x)dx \approx \epsilon f(a) \text{ khi } \epsilon \text{ rất nhỏ.}$$

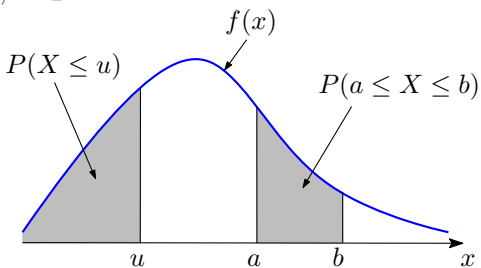
- Tập số thực $\{x \in \mathbb{R} : f(x) > 0\}$ được gọi là tập hỗ trợ (support) của $X$, kí hiệu $\text{Sup}(X)$.
- Hàm mật độ xác suất có tính chất: $f(x) \geq 0, \forall x \in \mathbb{R}$ và $\int_{-\infty}^{\infty} f(x)dx = 1$.

# Hàm mật độ xác suất

Hàm mật độ xác suất xác định phân phối của biến ngẫu nhiên liên tục

$$P(X \in C) = \int_C f(x)dx, C \subset \mathbb{R}.$$

- $P(X = u) = \int_u^u f(x)dx = 0$,
- $P(X < u) = P(X \leq u) = \int_{-\infty}^u f(x)dx$,
- $P(X > u) = P(X \geq u) = \int_u^{\infty} f(x)dx$,
- $P(a \leq X \leq b) = \int_a^b f(x)dx$.




Từ nhận xét $P(X = u) = 0$, ta thấy tồn tại các biến cố dù có xác suất là 0 nhưng vẫn có thể xảy ra (có $E$ với $P(E) = 0$ nhưng $E \neq \varnothing$).

# Biến ngẫu nhiên liên tục - Ví dụ 1

Biến ngẫu nhiên liên tục $X$ được gọi là có **phân phối đều liên tục** (continuous uniform distribution) trên $[a, b]$, kí hiệu $X \sim \mathcal{U}(a, b)$, nếu $X$ có hàm mật độ xác suất

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b, \\ 0 & \text{khác.} \end{cases}$$

$X$ là kết quả của thí nghiệm "chọn ngẫu nhiên một điểm trong khoảng $[a, b]$".

*Ví dụ:* một môn học dài 2 giờ, giáo viên điểm danh ngẫu nhiên trong thời gian học, bạn đi trễ $t$ phút. Gọi $X$ là thời điểm giáo viên điểm danh thì $X \sim \mathcal{U}(0, 2)$ (giờ), xác suất bạn được điểm danh là

$$P\left(X \geq \frac{t}{60}\right) = \int_{t/60}^{\infty} f(x)dx = \int_{t/60}^{2} \frac{1}{2}dx = \frac{1}{2}\left(2 - \frac{t}{60}\right) = 1 - \frac{t}{120}, 0 \leq t \leq 120.$$

# Biến ngẫu nhiên liên tục - Ví dụ 2

Xét $X$ là biến ngẫu nhiên liên tục với hàm mật độ xác suất có dạng

$$f(x) = \begin{cases} cx & 0 \leq x \leq 4, \\ 0 & \text{khác.} \end{cases}$$

Để $f$ là hàm mật độ xác suất hợp lệ, ta có điều kiện cho hệ số $c$ là

$$\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_0^4 cxdx = 1 \Rightarrow c\frac{x^2}{2}\Big|_{x=0}^{x=4} = 8c = 1 \Rightarrow c = \frac{1}{8}.$$

Xác suất $X$ nhận giá trị lớn hơn 2 là

$$P(X > 2) = \int_2^{\infty} f(x)dx = \int_2^4 \frac{1}{8}xdx = \frac{3}{4}.$$

# Nội dung

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

2. Hàm phân phối

3. Kì vọng và phương sai

4. Một số phân phối xác suất liên tục điển hình

5. Phương pháp xấp xỉ phân phối bằng mô phỏng

# Hàm phân phối

**Hàm phân phối (tích lũy)** (distribution function, cumulative distribution function) của một biến ngẫu nhiên $X$ là hàm số $F : \mathbb{R} \to \mathbb{R}$ được xác định bởi

$$F(x) = P(X \leq x) = \begin{cases} \sum_{t \leq x} f(t) & \text{nếu } X \text{ rời rạc với hàm xác suất } f, \\ \int_{-\infty}^{x} f(t)dt & \text{nếu } X \text{ liên tục với hàm mật độ xác suất } f. \end{cases}$$

$F$ xác định phân phối của $X$.

Hàm phân phối $F$ có các tính chất

1. Tăng: nếu $x_1 \leq x_2$ thì $F(x_1) \leq F(x_2)$,
2. Chuẩn hóa: $lim_{x \to -\infty} F(x) = 0$ và $lim_{x \to \infty} F(x) = 1$,
3. Liên tục phải: $F(x) = F(x^+) = lim_{t \to x, t > x} F(t)$.
4. Nếu $X$ là biến ngẫu nhiên liên tục thì $F$ là hàm liên tục và nếu $F$ có đạo hàm tại $x$ thì $F'(x) = f(x)$.

# Hàm phân phối - Ví dụ (biến ngẫu nhiên rời rạc)

Xét thí nghiệm tung một đồng xu (đồng chất) 2 lần, gọi $X$ là số lần được mặt ngửa. Hàm xác suất và hàm phân phối của $X$ là

$$f(x) = \begin{cases} 1/4 & x = 0, \\ 2/4 & x = 1, \\ 1/4 & x = 2, \\ 0 & x \notin \{0, 1, 2\}, \end{cases} \quad \text{và} \quad F(x) = \begin{cases} 0 & x < 0, \\ 1/4 & 0 \le x < 1, \\ 3/4 & 1 \le x < 2, \\ 1 & 2 \le x. \end{cases}$$

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X = x)$ | 1/4 | 1/2 | 1/4 |
| $P(X \le x)$ | 1/4 | 3/4 | 1 |

# Hàm phân phối - Ví dụ (biến ngẫu nhiên liên tục)

Xét $X$ là biến ngẫu nhiên liên tục với hàm mật độ xác suất

$$f(x) = \begin{cases} x/8 & 0 \le x \le 4, \\ 0 & \text{khác.} \end{cases}$$

Hàm phân phối của $X$ là

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \int_{-\infty}^{x} 0dt = 0 & x < 0, \\ \int_{0}^{x} \frac{t}{8}dt = \frac{x^2}{16} & 0 \le x < 4, \\ \int_{0}^{4} \frac{t}{8}dt = 1 & 4 \le x. \end{cases}$$

Từ đó ta có

$$P(1 \le X \le 2) = P(X \le 2) - P(X < 1) = F(2) - F(1^-) = F(2) - F(1) = \frac{2^2}{16} - \frac{1^2}{16} = \frac{3}{16}.$$

# Hàm phân phối đồng thời

**Hàm phân phối đồng thời** (joint distribution function) của hai biến ngẫu nhiên $X, Y$ là hàm số $F_{XY} : \mathbb{R}^2 \to \mathbb{R}$ được xác định bởi

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) \ (x, y \in \mathbb{R}).$$

**Mệnh đề**. Hai biến ngẫu nhiên $X, Y$ độc lập khi và chỉ khi $F_{XY}(x, y) = F_X(x)F_Y(y)$ với mọi $x, y \in \mathbb{R}$.

Hai biến ngẫu nhiên $X, Y$ được gọi là **liên tục đồng thời** (jointly continuous) nếu có hàm số không âm $f_{XY} : \mathbb{R}^2 \to \mathbb{R}$ sao cho với mọi $C \in \mathbb{R}^2$ ta có

$$P((X, Y) \in C) = \iint_C f_{XY}(x, y)dxdy.$$

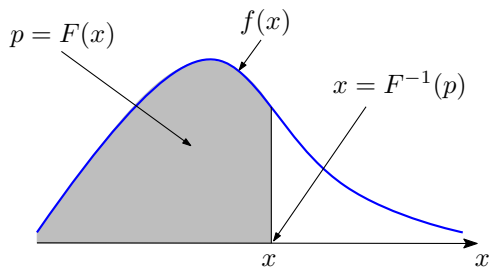
**Mệnh đề**. Hai biến ngẫu nhiên liên tục đồng thời $X, Y$ độc lập khi và chỉ khi $f_{XY}(x, y) = f_X(x)f_Y(y)$ với mọi $x, y \in \mathbb{R}$.

# Hàm phân vị

Cho biến ngẫu nhiên $X$ có hàm phân phối $F$, **hàm phân vị** (quantile function) của $X$ là hàm $Q : (0, 1) \to \mathbb{R}$, được xác định bởi

$$Q(p) = \text{“giá trị thực } x \text{ nhỏ nhất sao cho } F(x) \geq p\text{”}.$$

- $Q(p)$ được gọi là phân vị mức $p$ của phân phối của $X$ và thường được kí hiệu là $F^{-1}(p)$.
- Nếu $F$ liên tục và tăng ngặt thì $Q$ chính là hàm ngược $F^{-1}$ của $F$.
- Hàm phân vị $Q$ cho biết các điểm chia phân phối của $X$: $Q(50\%)$ được gọi là **trung vị** (median).

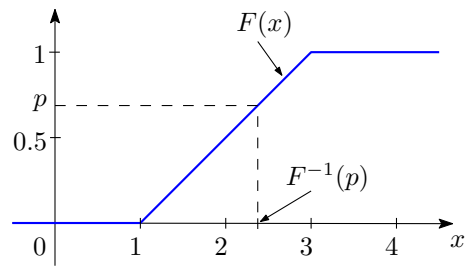


# Hàm phân vị - Ví dụ

Xét biến ngẫu nhiên $X \sim \mathcal{U}(1,3)$ có hàm mật độ $f(x) = \begin{cases} 1/2 & 1 \le x \le 3, \\ 0 & \text{khác.} \end{cases}$

$X$ có hàm phân phối

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0dt = 0 & x < 1, \\ \int_{1}^{x} \frac{1}{2}dt = \frac{x-1}{2} & 1 \le x < 3, \\ \int_{1}^{3} \frac{1}{2}dt = 1 & 3 \le x. \end{cases}$$




$X$ có hàm phân vị

$$Q(p) = x \Leftrightarrow \frac{x-1}{2} = p \Leftrightarrow x = 2p + 1$$

$$\Rightarrow Q(p) = F^{-1}(p) = 2p + 1, 0 < p < 1.$$

# Nội dung

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

2. Hàm phân phối

3. Kì vọng và phương sai

4. Một số phân phối xác suất liên tục điển hình

5. Phương pháp xấp xỉ phân phối bằng mô phỏng

# Kì vọng và phương sai

Cho biến ngẫu nhiên liên tục $X$ với hàm mật độ xác suất $f$

- **Kì vọng** (mean) của $X$ được tính bởi

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

- **Phương sai** (variance) của $X$ được tính bởi

$$\sigma^2 = Var(X) = E\left((X-\mu)^2\right) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx,$$

- Với hàm số $r : \mathbb{R} \to \mathbb{R}$ và $Y = r(X)$

$$E(Y) = E(r(X)) = \int_{-\infty}^{\infty} r(x)f(x)dx.$$

# Kì vọng và phương sai - Ví dụ 1

Xét $X$ là biến ngẫu nhiên liên tục với hàm mật độ xác suất

$$f(x) = \begin{cases} x/8 & 0 \le x \le 4, \\ 0 & \text{khác.} \end{cases}$$

$X$ có kì vọng

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^4 x\frac{x}{8}dx = \frac{x^3}{24}\Big|_{x=0}^{x=4} = \frac{8}{3}.$$

Hơn nữa

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^4 x^2\frac{x}{8}dx = 8,$$

nên $X$ có phương sai $Var(X) = E(X^2) - \mu^2 = \frac{8}{9}$ và độ lệch chuẩn $\sigma = \sqrt{Var(X)} = \frac{2\sqrt{2}}{3}$.

# Kì vọng và phương sai - Ví dụ 2

Xét các biến ngẫu nhiên liên tục $X_1, X_2, X_3$ có cùng tập hỗ trợ $[-1, 1]$ với các hàm mật độ xác suất lần lượt là

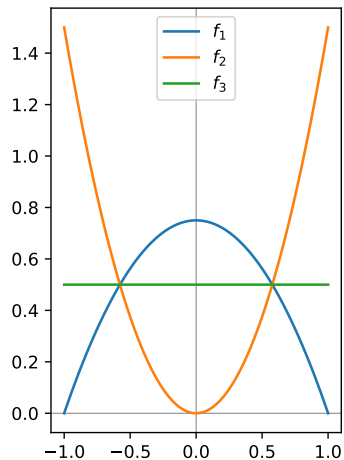
$$f_1(x) = -\frac{3x^2}{4} + \frac{3}{4}, \quad f_2(x) = \frac{3x^2}{2}, \quad f_3(x) = \frac{1}{2}.$$

Kì vọng của các biến

$$\mu_1 = \int_{-1}^{1} x\left(-\frac{3x^2}{4} + \frac{3}{4}\right) dx = 0, \quad \mu_2 = 0, \quad \mu_3 = 0.$$

Phương sai của các biến

$$\sigma_1^2 = \int_{-1}^{1} x^2\left(-\frac{3x^2}{4} + \frac{3}{4}\right) dx = \frac{1}{5}, \quad \sigma_2^2 = \frac{3}{5}, \quad \sigma_3^2 = \frac{1}{3}.$$



# Kì vọng và phương sai - Ví dụ 3

Chọn ngẫu nhiên một hình vuông có chiều dài trong khoảng $[0, 1]$ mét (m). Tính kì vọng diện tích hình vuông được chọn?

*Giải.* Gọi $X$ là chiều dài của hình vuông được chọn thì $X \sim \mathcal{U}(0, 1)$ nên $X$ có hàm mật độ xác suất

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1, \\ 0 & \text{khác.} \end{cases}$$

Gọi $Y$ là diện tích của hình vuông được chọn thì $Y = r(X)$ với hàm số $r(x) = x^2$. Từ đó ta có kì vọng của $Y$

$$E(Y) = E(r(X)) = \int_{\infty}^{\infty} r(x)f(x)dx = \int_0^1 x^2 1 dx = \frac{x^3}{3}\Big|_{x=0}^{x=1} = \frac{1}{3}m^2.$$

*Lưu ý:* ta dễ bị trực giác lừa là $E(S) = \frac{1}{2}m^2$. ($Y$ là biến ngẫu nhiên liên tục trên $[0, 1]$ nhưng không có phân phối đều.)

# Các tính chất quan trọng của kì vọng và phương sai

Cho $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên (có kỳ vọng), ta có

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) \quad (\textbf{linearity of expectation})$$

Cho $X$ là biến ngẫu nhiên và $a, b$ là các hằng số thực, ta có

1. $E(aX + b) = aE(X) + b$,
2. $Var(aX + b) = a^2 Var(X)$.

Cho $X, Y$ là hai biến ngẫu nhiên độc lập, ta có

1. $E(XY) = E(X)E(Y)$,
2. $Var(X + Y) = Var(X) + Var(Y)$.

# Nội dung

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

2. Hàm phân phối

3. Kì vọng và phương sai

4. Một số phân phối xác suất liên tục điển hình

5. Phương pháp xấp xỉ phân phối bằng mô phỏng

# Phân phối đều liên tục

Biến ngẫu nhiên liên tục $X$ được gọi là có **phân phối đều liên tục** (continuous uniform distribution) trên $[a, b]$ với $a < b$, kí hiệu $X \sim \mathcal{U}(a, b)$, nếu $X$ có tập giá trị là $[a, b]$ và

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{nếu } a \leq x \leq b, \\ 0 & \text{khác.} \end{cases}$$

Khi đó, $X$ có kì vọng $E(X) = \frac{a+b}{2}$ và phương sai $Var(X) = \frac{(b-a)^2}{12}$.

Gọi $X$ là kết quả của thí nghiệm "chọn ngẫu nhiên một điểm trong khoảng $[a, b]$" thì $X \sim \mathcal{U}(a, b)$.

**Mệnh đề.** Cho $X \sim \mathcal{U}(a, b)$ và $d \in (a, b)$, phân phối của $X$ khi biết $X \leq d$ là phân phối đều trên $[a, d]$, thường kí hiệu $(X|X \leq d) \sim \mathcal{U}(a, d)$.

# Phân phối mũ

Biến ngẫu nhiên liên tục $X$ được gọi là có **phân phối mũ** (exponential distribution) với tham số $\lambda$ ($\lambda > 0$), kí hiệu $X \sim \text{Exp}(\lambda)$, nếu $X$ có tập giá trị là $[0, \infty)$ và

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu } x \geq 0, \\ 0 & \text{khác.} \end{cases}$$

Khi đó, $X$ có kì vọng $E(X) = \frac{1}{\lambda}$, phương sai $Var(X) = \frac{1}{\lambda^2}$ và hàm phân phối

$$F(x) = 1 - e^{-\lambda x}, x \geq 0.$$

Phân phối mũ có thể được xem như là "phiên bản liên tục" của phân phối hình học.

**Mệnh đề** (tính không nhớ - memoryless). Cho $X \sim \text{Exp}(\lambda)$, với mọi $t, s \geq 0$ ta có

$$P(X > t + s | X > s) = P(X > t).$$

# Phân phối mũ - Ví dụ

Có $n$ bóng đèn, giả sử tuổi thọ của mỗi bóng đèn có phân phối mũ với tham số $\lambda > 0$ và độc lập nhau. Tính kì vọng thời gian chờ cho đến khi một bóng đèn nào đó bị hỏng?

*Giải.* Gọi $X_i$ là tuổi thọ của bóng đèn thứ $i$ ($i = 1, ..., n$), ta có $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. Gọi $Y$ là thời gian chờ cho đến khi một bóng đèn nào đó bị hỏng thì $Y = \min\{X_1, ..., X_n\}$. Với mọi $t > 0$ ta có

$$\begin{aligned} P(Y > t) &= P(\min\{X_1, ..., X_n\} > t) = P(X_1 > t, ..., X_n > t) \\ &= P(X_1 > t)...P(X_n > t) = (1 - (1 - e^{-\lambda t}))...(1 - (1 - e^{-\lambda t})) = e^{-\lambda t}...e^{-\lambda t} \\ &= e^{-n\lambda t}. \end{aligned}$$

Như vậy $Y$ có hàm phân phối

$$F_Y(t) = P(Y \leq t) = 1 - P(Y > t) = 1 - e^{-n\lambda t}$$

nên $Y \sim \text{Exp}(n\lambda)$. Do đó $E(Y) = \frac{1}{n\lambda}$.

# Phân phối chuẩn

Biến ngẫu nhiên liên tục $X$ được gọi là có **phân phối chuẩn** (normal distribution) với trung bình $\mu$ và phương sai $\sigma^2$ ($\sigma > 0$), kí hiệu $X \sim \mathcal{N}(\mu, \sigma^2)$, nếu $X$ có hàm mật độ xác suất

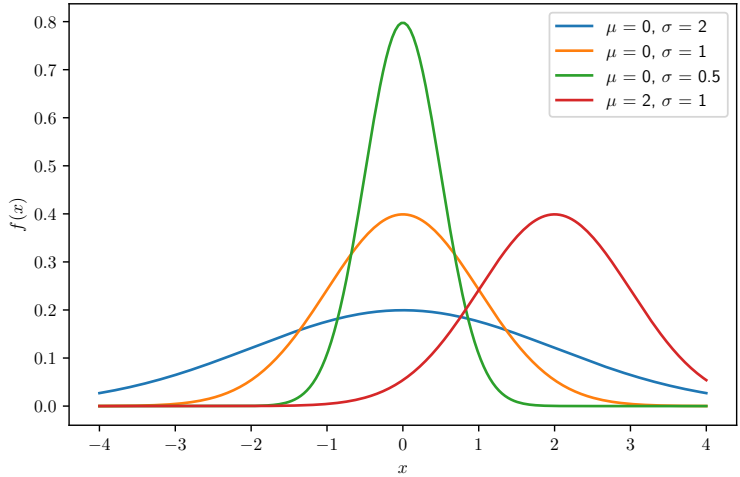
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}.$$

Khi đó, $X$ có kì vọng $E(X) = \mu$ và phương sai $Var(X) = \sigma^2$.

Trường hợp $Z \sim \mathcal{N}(0, 1)$ thì $Z$ được gọi là có **phân phối chuẩn tắc** (standard normal distribution). Hàm mật độ xác suất và hàm xác suất của $Z$ thường được kí hiệu lần lượt là $\phi, \Phi$, tức là

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

# Phân phối chuẩn - Hàm mật độ xác suất



# Phân phối chuẩn - Các tính chất quan trọng

1. Nếu $X \sim \mathcal{N}(\mu, \sigma^2)$ và $Y = aX + b$ $(a \neq 0)$ thì $X \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
2. Nếu $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ và $X_1, X_2$ độc lập thì $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
3. Nếu $Z \sim \mathcal{N}(0, 1)$ và $X = \sigma Z + \mu$ thì $X \sim \mathcal{N}(\mu, \sigma^2)$.
4. Nếu $X \sim \mathcal{N}(\mu, \sigma^2)$ và $Z = \frac{X-\mu}{\sigma}$ thì $Z \sim \mathcal{N}(0, 1)$, từ đó

$$F_X(x) = P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

5. Do tính đối xứng của $\phi$ nên
   - $\Phi(0) = 0.5, \Phi^{-1}(0.5) = 0$,
   - Với mọi $z \in \mathbb{R}$ và mọi $0 < p < 1$: $\Phi(-z) = 1 - \Phi(z), \Phi^{-1}(p) = -\Phi^{-1}(1-p)$,
   - Với mọi $z \geq 0$: $P(-z \leq Z \leq z) = 2\Phi(z) - 1$.

# Phân phối chuẩn - Ví dụ

Xét thí nghiệm bắt cá trong hồ, gọi $X$ là chiều dài cá bắt được. Giả sử $X$ có phân phối chuẩn với trung bình $\mu = 16cm$ và độ lệch chuẩn $\sigma = 4cm$. Đặt $Z = \frac{X-\mu}{\sigma} = \frac{X-16}{4}$.

- Xác suất bắt được cá nhỏ hơn 10cm

$$P(X < 10) = P\left(Z < \frac{10-16}{4}\right) = \Phi(-1.5) = 1 - \Phi(1.5) = 0.0668.$$

- Xác suất bắt được cá lớn hơn 24cm

$$P(X > 24) = 1 - P(X \leq 24) = 1 - P\left(Z \leq \frac{24-16}{4}\right) = 1 - \Phi(2) = 0.0228.$$

- 10% cá nhỏ nhất trong hồ có chiều dài $\leq a$ $(cm)$ là bao nhiêu?

$$P(X \leq a) = P\left(Z \leq \frac{a-16}{4}\right) = 0.1 \Rightarrow \frac{a-16}{4} = \Phi^{-1}(0.1) = -1.2816 \Rightarrow a = 10.87.$$

# Nội dung

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

2. Hàm phân phối

3. Kì vọng và phương sai

4. Một số phân phối xác suất liên tục điển hình

5. Phương pháp xấp xỉ phân phối bằng mô phỏng

# Phương pháp xấp xỉ kỳ vọng bằng mô phỏng

Để xấp xỉ kì vọng $E(X)$ của một biến ngẫu nhiên $X$ liên quan đến thí nghiệm $T$, ta có thể dùng phương pháp thống kê như sau

- Thực hiện lặp lại $N$ lần (độc lập) thí nghiệm $T$, ghi nhận các giá trị mà $X$ nhận $x_1, x_2, ..., x_N$ (còn gọi là **mẫu dữ liệu** - sample), và tính **trung bình mẫu**

$$\bar{x} = \frac{x_1 + x_2 + ... + x_N}{N}.$$

- Khi $N$ đủ lớn, ta có $\bar{x} \approx E(X)$.
- Việc thực hiện lặp lại nhiều lần thí nghiệm $T$ có thể được **mô phỏng** (simulate) trên máy tính.

# Phương pháp xấp xỉ phân phối bằng mô phỏng
# Biến ngẫu nhiên liên tục

Để xấp xỉ hàm mật độ xác suất $f_X$ của một biến ngẫu nhiên liên tục $X$ liên quan đến thí nghiệm $T$, ta có thể dùng phương pháp thống kê như sau

- Thực hiện lặp lại $N$ lần (độc lập) thí nghiệm $T$, ghi nhận các giá trị mà $X$ nhận $x_1, x_2, ..., x_N$ (còn gọi là **mẫu dữ liệu** - sample).
- Khi $N$ đủ lớn, ta có thể dùng **histogram** hoặc **ước lượng mật độ nhân** (kernel density estimation, KDE) trên mẫu để xấp xỉ $f_X$.
- Việc thực hiện lặp lại nhiều lần thí nghiệm $T$ có thể được **mô phỏng** (simulate) trên máy tính.

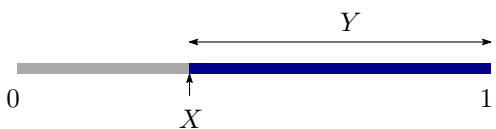
# Ví dụ 1

**Bài toán.** Chọn ngẫu nhiên một điểm trên một thanh có chiều dài 1 đơn vị, cắt tại điểm đó thành hai đoạn và giữ lại đoạn dài hơn. Tính kì vọng và tìm phân phối của chiều dài đoạn giữ lại.

*Giải.* Gọi $X$ là vị trí ngẫu nhiên chọn trên thanh thì $X \sim \mathcal{U}(0,1)$. Do đó $X$ là biến ngẫu nhiên liên tục với hàm mật độ xác suất

$$f_X(x) = \begin{cases} 1 & \text{nếu } 0 \le x \le 1, \\ 0 & \text{khác.} \end{cases}$$

Gọi $Y$ là chiều dài của đoạn được giữ lại (tức là đoạn dài hơn) thì $Y = \max\{X, 1-X\}$.



# Ví dụ 1 (tt)

Ta có kì vọng của chiều dài đoạn giữ lại

$$E(Y) = E\left(\max\{X, 1-X\}\right) = \int_{-\infty}^{\infty} \max\{x, 1-x\} f_X(x) dx = \int_0^1 \max\{x, 1-x\} dx$$
$$= \int_0^{1/2} \max\{x, 1-x\} dx + \int_{1/2}^1 \max\{x, 1-x\} dx = \int_0^{1/2} (1-x) dx + \int_{1/2}^1 x dx$$
$$= \frac{3}{4}.$$

# Ví dụ 1 (tt)

Ta tìm hàm phân phối (tích lũy) của biến ngẫu nhiên $Y = \max\{X, 1-X\}$

$$F_Y(y) = P(Y \leq y) = P(\max\{X, 1-X\} \leq y), y \in \mathbb{R}.$$

Xét các trường hợp của $y$

1. $y < 1/2$: $(\max\{X, 1-X\} \leq y) = \varnothing$ vì $0 \leq X \leq 1$ nên $1/2 \leq \max\{X, 1-X\}$,

$$P\left(\max\{X, 1-X\} \leq y\right) = P(\varnothing) = 0.$$

2. $1/2 \leq y \leq 1$: $(\max\{X, 1-X\} \leq y) = (1-y \leq X \leq y)$,

$$P\left(\max\{X, 1-X\} \leq y\right) = P\left(1-y \leq X \leq y\right) = \int_{1-y}^{y} f_X(x)dx = \int_{1-y}^{y} dy = 2y.$$

3. $y > 1$: $(\max\{X, 1-X\} \leq y) = \Omega$ vì $0 \leq X \leq 1$ nên $\max\{X, 1-X\} \leq 1$,

$$P\left(\max\{X, 1-X\} \leq y\right) = P(\Omega) = 1.$$

# Ví dụ 1 (tt)

Từ đó ta có

$$F_Y(y) = \begin{cases} 0 & \text{nếu } y < 1/2, \\ 2y & \text{nếu } 1/2 \leq y \leq 1, \\ 1 & \text{nếu } 1 < y. \end{cases}$$

Lấy đạo hàm của hàm phân phối, ta có hàm mật độ xác suất của $Y$ là

$$f_Y(y) = F'_Y(y) = \begin{cases} 2 & \text{nếu } 1/2 \leq x \leq 1, \\ 0 & \text{khác.} \end{cases}$$

Như vậy $Y$ có phân phối đều trên đoạn $[1/2, 1]$, tức là $Y \sim \mathcal{U}(1/2, 1)$.
Lưu ý, từ phân phối của $Y$, $Y \sim \mathcal{U}(1/2, 1)$, ta cũng có $E(Y) = \frac{1/2+1}{2} = \frac{3}{4}$.

# Ví dụ 1 (tt)

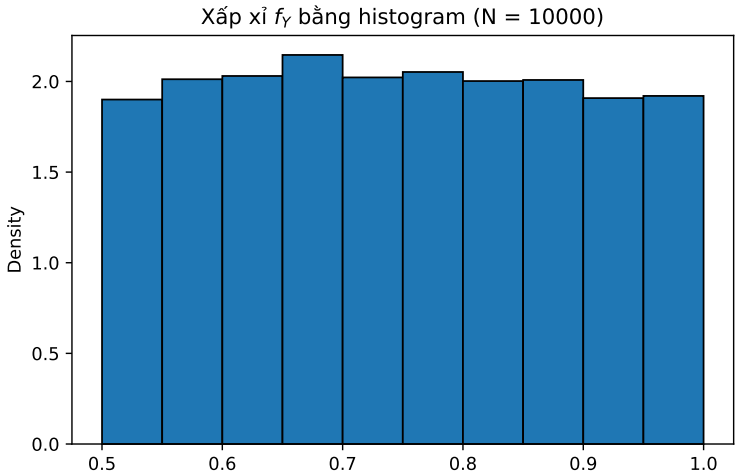
```
def greater_len(N):
    X = np.random.uniform(size=N)
    Y = np.maximum(X, 1 - X)
    return Y

N = 10000

np.mean(greater_len(N))
#0.7499721269808018

plt.hist(greater_len(N), density=True, edgecolor="black")
```

# Ví dụ 1 (tt)


Xấp xỉ $f_Y$ bằng histogram (N = 10000)
Density
0.0
0.5
1.0
1.5
2.0
0.5
0.6
0.7
0.8
0.9
1.0

# Ví dụ 2

**Bài toán.** Cho $X_1, X_2, ..., X_n$ là $n$ biến ngẫu nhiên độc lập và cùng phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$. Đặt

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{và} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

($X_1, ..., X_n$ thường được gọi là một mẫu ngẫu nhiên cỡ $n$, $\bar{X}$ là trung bình mẫu và $S^2$ là phương sai mẫu.)
Tìm phân phối của các biến ngẫu nhiên $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ và $\frac{\bar{X}-\mu}{S/\sqrt{n}}$.

*Trả lời:* $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ có phân phối chuẩn tắc $\mathcal{N}(0,1)$ và $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ có **phân phối Student** (Student's t-distribution) với $n-1$ bậc tự do.
(`https://en.wikipedia.org/wiki/Student%27s_t-distribution`.)

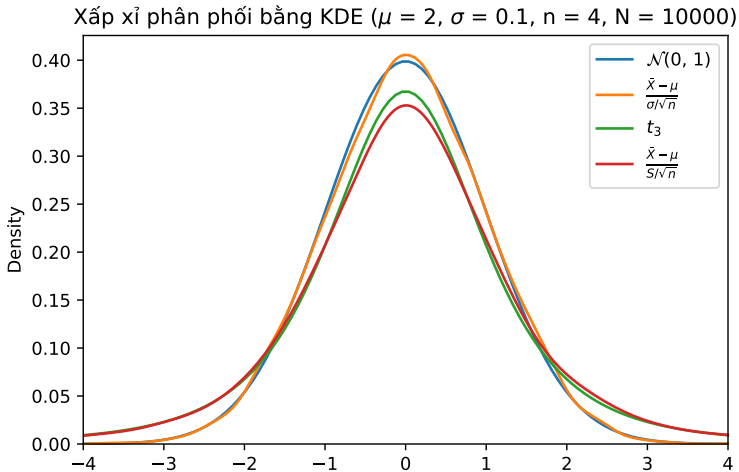
# Ví dụ 2 (tt)

```
def sample(mu, sigma, n, N):
    X = np.random.normal(mu, sigma, size=(N, n))
    X_bar = np.mean(X, axis=1)
    S2 = np.var(X, axis=1, ddof=1)
    return X_bar, S2

X_bar, S2 = sample(mu, sigma, n, N)

plt.plot(x, scipy.stats.norm.pdf(x))
sns.kdeplot((X_bar - mu)/(sigma/np.sqrt(n)))
plt.plot(x, scipy.stats.t.pdf(x, n - 1))
sns.kdeplot((X_bar - mu)/(np.sqrt(S2)/np.sqrt(n)))
```

# Ví dụ 2 (tt)


Xấp xỉ phân phối bằng KDE ($\mu$ = 2, $\sigma$ = 0.1, n = 4, N = 10000)
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
−4
−3
−2
−1
0
1
2
3
4
$\mathcal{N}(0, 1)$
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
$t_3$
$\frac{\bar{X}-\mu}{S/\sqrt{n}}$

# Tài liệu tham khảo

**Chapter 3-5.** Morris H. DeGroot, Mark J. Schervish. *Probability and Statistics*. Addison-Wesley, 2012.

**Chapter 4.** H. Pishro-Nik. *"Introduction to probability, statistics, and random processes"*, available at `https://www.probabilitycourse.com`. Kappa Research LLC, 2014.